

Lazy TSO Reachability

Ahmed Bouajjani¹, Georgel Calin², Egor Derevenetc^{2,3}, and Roland Meyer²

¹LIAFA, University Paris 7 ²University of Kaiserslautern ³Fraunhofer ITWM

Abstract. We address the problem of checking state reachability for programs running under Total Store Order (TSO). The problem has been shown to be decidable but the cost is prohibitive, namely non-primitive recursive. We propose here to give up completeness. Our contribution is a new algorithm for TSO reachability: it uses the standard SC semantics and introduces the TSO semantics lazily and only where needed. At the heart of our algorithm is an iterative refinement of the program of interest. If the program's goal state is SC-reachable, we are done. If the goal state is not SC-reachable, this may be due to the fact that SC under-approximates TSO. We employ a second algorithm that determines TSO computations which are infeasible under SC, and hence likely to lead to new states. We enrich the program to emulate, under SC, these TSO computations. Altogether, this yields an iterative under-approximation that we prove sound and complete for bug hunting, i.e., a semi-decision procedure halting for positive cases of reachability. We have implemented the procedure as an extension to the tool Trencher [1] and compared it to the MEMORAX [2] and CBMC [14] model checkers.

1 Introduction

Sequential consistency (SC) [21] is the semantics typically assumed for parallel programs. Under SC, instructions are executed atomically and in program order. When programs are executed on an Intel x86 processor, however, they are only guaranteed a weaker semantics known as Total Store Order (TSO). TSO weakens the synchronization guarantees given by SC, which in turn may lead to erroneous behavior. TSO reflects the architectural optimization of store buffers. To reduce the latency of memory accesses, store commands are added to a thread-local FIFO buffer and only later executed on memory.

To check for correct behavior, reachability techniques have proven useful. Given a program and a goal state, the task is to check whether the state is reachable. To give an example, assertion failures can be phrased as reachability problems. Reachability depends on the underlying semantics. Under SC, the problem is known to be PSPACE-complete [18]. Under TSO, it is considerably more difficult: although decidable, it is non-primitive recursive-hard [8].

Due to the high complexity, tools rarely provide decision procedures [2, 23, 24]. Instead, most approaches implement approximations. Typical approximations of TSO reachability bound the number of loop iterations [5, 6], the number of context switches between threads [9], or the size of store buffers [19, 20]. What

all these approaches have in common is that they introduce store buffering in the *whole* program. We claim that such a comprehensive instrumentation is unnecessarily heavy.

The idea of our method is to introduce store buffering lazily and only where needed. Unlike [2], we do not target completeness. Instead, we argue that our lazy TSO reachability checker is useful for a fast detection of bugs that are due to the TSO semantics. At a high level, we solve the expensive TSO reachability problem with a series of cheap SC reachability checks — very much like SAT solvers are invoked as subroutines of costlier analyses. The SC checks run interleaved with queries to an oracle. The task of the oracle is to suggest sequences of instructions that should be considered under TSO, which means they are likely to lead to TSO-reachable states outside SC.

To be more precise, the algorithm iteratively repeats the following steps. First, it checks whether the goal state is SC-reachable. If this is the case, the state will be TSO-reachable as well and the algorithm returns. If the state is not SC-reachable, the algorithm asks the oracle for a sequence of instructions and encodes the TSO behavior of the sequence into the input program. As a result, precisely this TSO behavior becomes available under SC. The encoding is linear in the size of the input program and in the length of the sequence.

The algorithm is a semi-decision procedure: it always returns correct answers and is guaranteed to terminate if the goal state is TSO-reachable. This guarantee relies on one assumption on the oracle. If the oracle returns the empty sequence, then the SC- and the TSO-reachable states of the input program have to coincide. We also come up with a good oracle: robustness checkers naturally meet the above requirement. Intuitively, a program is robust against TSO if its partial order-behaviors (reflecting data and control dependencies) under TSO and under SC coincide. Robustness is much easier than TSO reachability, actually PSPACE-complete [10,11], and hence well-suited for iterative invocations.

We have implemented lazy TSO reachability as an extension to our tool TRENCHER [1], reusing the robustness checking algorithms of TRENCHER to derive an oracle. The implementation is able to solve positive instances of TSO reachability as well as correctly determine safety for robust programs. The source code and experiments are available online [1].

The structure of the paper is as follows. We introduce parallel programs with their TSO and their SC semantics in Section 2. Section 3 presents our main contribution, the lazy approach to solving TSO reachability. Section 4 describes the robustness-based oracle. The experimental evaluation is given in Section 5. Details and proofs missing in the main text can be found in the appendix.

Related Work

As already mentioned, TSO reachability was proven decidable but non-primitive recursive [8] in the case of a finite number of threads and a finite data domain. In the same setting, robustness was shown to be PSPACE-complete [11]. Checking and enforcing robustness against weak memory models has been addressed in

[3, 7, 10–13, 26]. The first work to give an efficient sound and complete decision procedure for checking robustness is [10].

The works [2, 23, 24] propose state-based techniques to solve TSO reachability. An under-approximative method that uses bounded context switching is given in [9]. It encodes store buffers into a linear-size instrumentation, and the instrumented program is checked for SC reachability. The under-approximative techniques of [5, 6] are able to guarantee safety only for programs with bounded loops. On the other side of the spectrum, over-approximative analyses abstract store buffers into sets combined with bounded queues [19, 20].

2 Parallel Programs

We use automata to define the syntax and the semantics of parallel programs. A (non-deterministic) *automaton* over an alphabet Σ is a tuple $A = (\Sigma, S, \rightarrow, s_0)$, where S is a set of states, $\rightarrow \subseteq S \times (\Sigma \cup \{\varepsilon\}) \times S$ is a set of transitions, and $s_0 \in S$ is an initial state. The automaton is *finite* if the transition relation \rightarrow is finite. We write $s \xrightarrow{a} s'$ if $(s, a, s') \in \rightarrow$, and extend the transition relation to sequences $w \in \Sigma^*$ as expected. The *language of A with final states $F \subseteq S$* is $\mathcal{L}_F(A) := \{w \in \Sigma^* \mid s_0 \xrightarrow{w} s \in F\}$. We say that state $s \in S$ is *reachable* if $s_0 \xrightarrow{w} s$ for some sequence $w \in \Sigma^*$. Letter a *precedes* b in w , denoted by $a <_w b$, if $w = w_1 \cdot a \cdot w_2 \cdot b \cdot w_3$ for some $w_1, w_2, w_3 \in \Sigma^*$.

A parallel program P is a finite sequence of threads that are identified by indices t from TID. Each thread $t := (Com_t, Q_t, I_t, q_{0,t})$ is a finite automaton with transitions I_t that we call *instructions*. Instructions I_t are labelled by *commands* from the set Com_t which we define in the next paragraph. We assume, wlog., that states of different threads are disjoint. This implies that the sets of instructions of different threads are distinct. We use $I := \biguplus_{t \in \text{TID}} I_t$ for all instructions and $Com := \bigcup_{t \in \text{TID}} Com_t$ for all commands. For an instruction $inst := (s, cmd, s')$ in I , we define $cmd(inst) := cmd$, $src(inst) := s$, and $dst(inst) := s'$.

To define the set of commands, let DOM be a finite domain of values that we also use as addresses. We assume that value 0 is in DOM. For each thread t , let REG_t be a finite set of registers that take their values from DOM. We assume per-thread disjoint sets of registers. The set of expressions of thread t , denoted by EXP_t , is defined over registers from REG_t , constants from DOM, and (unspecified) operators over DOM. If $r \in REG_t$ and $e, e' \in EXP_t$, the set of commands Com_t consists of loads from memory $r \leftarrow mem[e]$, stores to memory $mem[e] \leftarrow e'$, memory fences **mfence**, assignments $r \leftarrow e$, and conditionals **assume** e . We write $REG := \biguplus_{t \in \text{TID}} REG_t$ for all registers and $EXP := \bigcup_{t \in \text{TID}} EXP_t$ for all expressions.

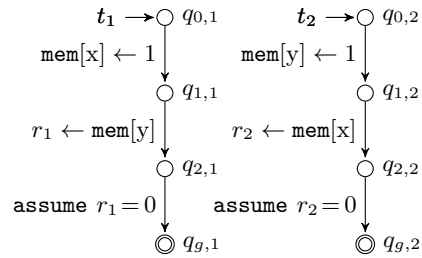


Fig. 1. Simplified Dekker's algorithm.

The program in Figure 1 serves as our running example. It consists of two threads t_1 and t_2 implementing a mutual exclusion protocol. Initially, the addresses x and y contain 0. The first thread signals its intent to enter the critical section by setting variable x to 1. Next, the thread checks whether the second thread wants to enter the critical section, too. It reads variable y and, if it is 0, the first thread enters its critical section. The critical section actually is the state $q_{g,1}$. The second thread behaves symmetrically.

2.1 Semantics of Parallel Programs

The semantics of a parallel program P under memory model $M = \text{TSO}$ and $M = \text{SC}$ follows [25]. We define the semantics in terms of a *state-space automaton* $X_M(P) := (E, S_M, \Delta_M, s_0)$. Each state $s = (\text{pc}, \text{val}, \text{buf}) \in S_M$ is a tuple where the program counter $\text{pc}: \text{TID} \rightarrow Q$ holds the current control state of each thread, the valuation $\text{val}: \text{REG} \cup \text{DOM} \rightarrow \text{DOM}$ holds the values stored in registers and at memory addresses, and the buffer configuration $\text{buf}: \text{TID} \rightarrow (\text{DOM} \times \text{DOM})^*$ holds a sequence of address-value pairs.

In the *initial state* $s_0 := (\text{pc}_0, \text{val}_0, \text{buf}_0)$, the program counter holds the initial control states, $\text{pc}_0(t) := q_{0,t}$ for all $t \in \text{TID}$, all registers and addresses contain value 0, and all buffers are empty, $\text{buf}_0(t) := \varepsilon$ for all $t \in \text{TID}$.

The transition relation Δ_{TSO} for TSO satisfies the rules given in Figure 2. There are two more rules for register assignments and conditionals that are standard and omitted. TSO architectures implement (FIFO) store buffering, which means stores are buffered for later execution on the shared memory. Loads from an address a take their value from the most recent store to address a that is buffered. If there is no such buffered store, they access the main memory. This is modelled by the Rules (LB) and (LM). Rule (ST) enqueues store operations as address-value pairs to the buffer. Rule (MEM) non-deterministically dequeues store operations and executes them on memory. Rule (F) states that a thread can execute a fence only if its buffer is empty. As can be seen from Figure 2, events labelling TSO transitions take the form $E \subseteq \text{TID} \times (I \cup \{\text{flush}\}) \times (\text{DOM} \cup \{\perp\})$.

The SC [21] semantics is simpler than TSO in that stores are not buffered. Technically, we keep the set of states but change the transitions so that Rule (ST) is immediately followed by Rule (MEM).

We are interested in the *computations* of program P under $M \in \{\text{TSO}, \text{SC}\}$. They are given by $\mathcal{C}_M(P) := \mathcal{L}_F(X_M(P))$, where F is the set of states with empty buffers. With this choice of final states, we avoid incomplete computations that have pending stores. Note that all SC states have empty buffers, which means the SC computations form a subset of the TSO computations: $\mathcal{C}_{\text{SC}}(P) \subseteq \mathcal{C}_{\text{TSO}}(P)$. We will use notation $\text{Reach}_M(P)$ for the set of all states $s \in F$ that are reachable by some computation in $\mathcal{C}_M(P)$.

To give an example, the program from Figure 1 admits the TSO computation τ_{wit} below where the store of the first thread is flushed at the end:

$$\tau_{\text{wit}} = \text{store}_1 \cdot \text{load}_1 \cdot \text{store}_2 \cdot \text{flush}_2 \cdot \text{load}_2 \cdot \text{flush}_1.$$

$$\begin{array}{c}
 \frac{cmd = r \leftarrow \text{mem}[e_a] \quad \text{buf}(t) \downarrow (\{a\} \times \text{DOM}) = (a, v) \cdot \beta}{s \xrightarrow{(t, \text{inst}, a)} (\text{pc}', \text{val}[r := v], \text{buf})} \quad (\text{LB}) \\
 \\
 \frac{cmd = r \leftarrow \text{mem}[e_a] \quad \text{buf}(t) \downarrow (\{a\} \times \text{DOM}) = \varepsilon}{s \xrightarrow{(t, \text{inst}, a)} (\text{pc}', \text{val}[r := \text{val}(a)], \text{buf})} \quad (\text{LM}) \\
 \\
 \frac{cmd = \text{mem}[e_a] \leftarrow e_v}{s \xrightarrow{(t, \text{inst}, a)} (\text{pc}', \text{val}, \text{buf}[t := (a, v) \cdot \text{buf}(t)])} \quad (\text{ST}) \\
 \\
 \frac{\text{buf}(t) = \beta \cdot (a, v)}{s \xrightarrow{(t, \text{flush}, a)} (\text{pc}, \text{val}[a := v], \text{buf}[t := \beta])} \quad (\text{MEM}) \qquad \frac{cmd = \text{mfence} \quad \text{buf}(t) = \varepsilon}{s \xrightarrow{(t, \text{inst}, \perp)} (\text{pc}', \text{val}, \text{buf})} \quad (\text{F})
 \end{array}$$

Fig. 2. Transition rules for $X_{\text{TSO}}(P)$ assuming $s = (\text{pc}, \text{val}, \text{buf})$ with $\text{pc}(t) = q$ and $\text{inst} = (q, cmd, q')$ in thread t . The program counter is always set to $\text{pc}' = \text{pc}[t := q']$. We assume $a = \hat{e}_a$ to be the address returned by an address expression e_a and $v = \hat{e}_v$ the value returned by a value expression e_v . We use $\text{buf}(t) \downarrow (\{a\} \times \text{DOM})$ to project the buffer content $\text{buf}(t)$ to store operations that access address a .

Consider an event $\mathbf{e} = (t, \text{inst}, a)$. By $\text{thread}(\mathbf{e}) := t$ we refer to the thread that produced the event. Function $\text{inst}(\mathbf{e}) := \text{inst}$ returns the instruction. For flush events, $\text{inst}(\mathbf{e})$ gives the instruction of the matching store event. By $\text{addr}(\mathbf{e}) := a$ we denote the address that is accessed (if any). In the example, $\text{thread}(\text{store}_1) = t_1$, $\text{inst}(\text{store}_1) = q_{0,1} \xrightarrow{\text{mem}[x] \leftarrow 1} q_{1,1}$, and $\text{addr}(\text{store}_1) = x$.

3 Lazy TSO Reachability

We introduce the reachability problem and present our main contribution: an algorithm that checks TSO reachability lazily. The iterative algorithm queries an oracle to identify sequences of instructions that, under the TSO semantics, lead to states not reachable under SC. In Section 3.1, we show that the algorithm yields a sound and complete semi-decision procedure.

Given a memory model $M \in \{\text{SC}, \text{TSO}\}$, the M reachability problem expects as input a program P and a set of *goal states* $G \subseteq S_M$. We are mostly interested in the control state of each thread. Therefore, goal states $(\text{pc}, \text{val}, \text{buf})$ typically specify a program counter pc but leave the memory valuation unconstrained. Formally, the M reachability problem asks if some state in G is reachable in the automaton $X_M(P)$.

Given: A parallel program P and goal states G .

Problem: Decide $\mathcal{L}_{F \cap G}(X_M(P)) \neq \emptyset$.

We use notation $\text{Reach}_M(P) \cap G$ for the set of reachable final goal states in P .

Instead of solving reachability under TSO directly, the algorithm we propose solves SC reachability and, if no goal state is reachable, tries to lazily introduce store buffering on a certain control path of the program. The algorithm delegates choosing the control path to an *oracle function* \mathcal{O} . Given an input program R ,

the oracle returns a sequence of instructions I^* in that program. Formally, the oracle satisfies the following requirements:

- If $\mathcal{O}(R) = \varepsilon$ then $\text{Reach}_{\text{SC}}(R) = \text{Reach}_{\text{TSO}}(R)$.
- Otherwise, $\mathcal{O}(R) = \text{inst}_1 \text{inst}_2 \dots \text{inst}_n$ with $\text{cmd}(\text{inst}_1)$ a store, $\text{cmd}(\text{inst}_n)$ a load, $\text{cmd}(\text{inst}_i) \neq \text{mfence}$, and $\text{dst}(\text{inst}_i) = \text{src}(\text{inst}_{i+1})$ for $i \in [1..n-1]$.

The lazy TSO reachability checker is outlined in Algorithm 1. As input, it takes a program P and an oracle \mathcal{O} . We assume some control states in each thread to be marked to define a set of goal states. The algorithm returns *true* iff the program can reach a goal state under TSO. It works as follows. First, it creates a copy R of the program P . Next, it checks if a goal state is SC-reachable in R (Line 3). If that is the case, the algorithm returns *true*. Otherwise, it asks the oracle \mathcal{O} where in the program to introduce store buffering. If $\mathcal{O}(R) \neq \varepsilon$, the algorithm extends R to emulate store buffering on the path $\mathcal{O}(R)$ under SC (Line 8). Then it goes back to the beginning of the loop. If $\mathcal{O}(R) = \varepsilon$, by the first property of oracles, R has the same reachable states under SC and under TSO. This means the algorithm can safely return *false* (Line 10). Note that, since R emulates TSO behavior of P , the algorithm solves TSO reachability for P .

Algorithm 1 Lazy TSO reachability Checker

Input: Marked program P and oracle \mathcal{O}

Output: *true* if some goal state is TSO-reachable in P
false if no goal state is TSO-reachable in P

```

1:  $R := P$ ;
2: while true do
3:   if  $\text{Reach}_{\text{SC}}(P) \cap G \neq \emptyset$  then    {check if some goal state is SC-reachable}
4:     return true;
5:   else
6:      $\sigma := \mathcal{O}(R)$ ;                        {ask the oracle where to use store buffering}
7:     if  $\sigma \neq \varepsilon$  then
8:        $R := R \oplus \sigma$ ;
9:     else
10:    return false;

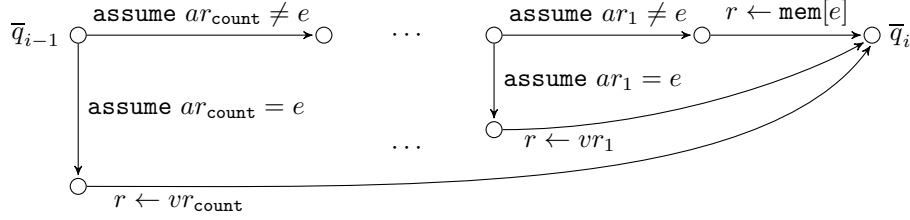
```

Let $\sigma := \mathcal{O}(R) = \text{inst}_1 \text{inst}_2 \dots \text{inst}_n$ and let $t := (\text{Com}_t, Q_t, I_t, q_{0,t})$ be the thread of the instructions in σ . The modified program $R \oplus \sigma$ replaces t by a new thread $t \oplus \sigma$. The new thread emulates under SC the TSO semantics of σ . Formally, the *extension of t by σ* is $t \oplus \sigma := (\text{Com}'_t, Q'_t, I'_t, q_{0,t})$. The thread is obtained from t by adding sequences of instructions starting from $\bar{q}_0 := \text{src}(\text{inst}_1)$. To remember the addresses and values of the buffered stores, we use auxiliary registers ar_1, \dots, ar_{\max} and vr_1, \dots, vr_{\max} , where $\max \leq n-1$ is the total number of store instructions in σ . The sets $\text{Com}'_t \supseteq \text{Com}_t$ and $Q'_t \supseteq Q_t$ are extended as necessary.

We define the extension by describing the new transitions that are added to I'_t for each $inst_i$. In our construction, we use a variable `count` to keep track of the number of store instructions already processed. Initially, $Q'_t := Q_t$ and `count` := 0. Based on the type of instructions, we distinguish the following cases.

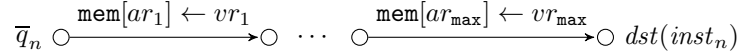
If $cmd(inst_i) = \text{mem}[e] \leftarrow e'$, we increment `count` by 1 and add instructions that remember the address and the value being written in ar_{count} and vr_{count} .

If $cmd(inst_i) = r \leftarrow \text{mem}[e]$, we add instructions to I'_t that perform a load from memory only when a load from the simulated buffer is not possible. More precisely, if $j \in [1, \text{count}]$ is found so that $ar_j = e$, register r is assigned the value of vr_j . Otherwise, r receives its value from the address indicated by e .

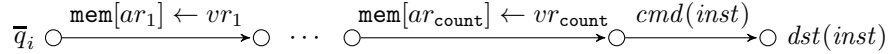


If $cmd(inst_i)$ is an assignment or a conditional, we add $(\bar{q}_{i-1}, cmd(inst_i), \bar{q}_i)$ to I'_t . By the definition of an oracle, $cmd(inst_i)$ is never a fence.

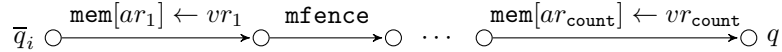
The above cases handle all instructions in σ . So far, the extension added new instructions to I'_t that lead through the fresh states $\bar{q}_1, \dots, \bar{q}_n$. Out of control state \bar{q}_n , we now recreate the sequence of stores remembered by the auxiliary registers. Then we return to the control flow of the original thread t .



Next, we remove $inst_1$ from the program. This prevents the oracle from discovering in the future another instruction sequence that is essentially the same as σ . As we will show, this is key to guaranteeing termination of the algorithm for acyclic programs. However, the removal of $inst_1$ may reduce the set of TSO-reachable states. To overcome this problem, we insert additional instructions. Consider an instruction $inst \in I_t$ with $src(inst) = src(inst_i)$ for some $i \in [1..n]$ and assume that $inst \neq inst_i$. We add instructions that recreate the stores buffered in the auxiliary registers and return to $dst(inst)$.



Similarly, for all load instructions $inst_i$ as well as out of \bar{q}_1 we add instructions that flush and fence the pair (ar_1, vr_1) , make visible the remaining buffered stores, and return to state q in the original control flow. Below, $q := src(inst_i)$ if $inst_i$ is a load and $q := dst(inst_1)$, otherwise. Intuitively, this captures behaviors that delay $inst_1$ past loads earlier than $inst_n$, and that do not delay $inst_1$ past the first load in σ .



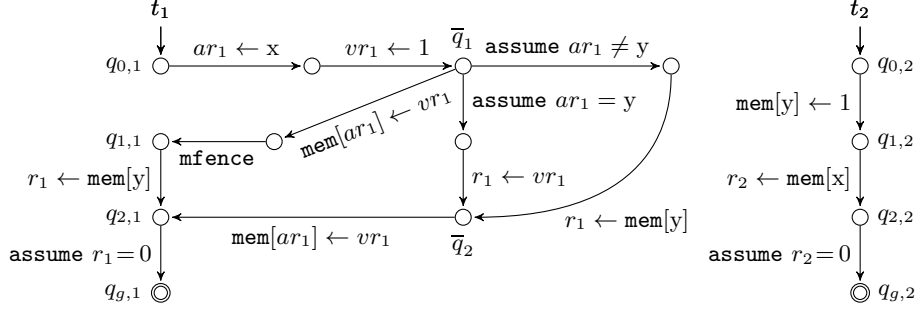


Fig. 3. Extension by $inst(store_1) \cdot inst(load_1)$ of the program in Figure 1. Goal state (pc, val, buf) with $val(x) = val(y) = 1$ and $val(r_1) = val(r_2) = 0$ is now SC-reachable.

Figure 3 shows the extension of the program in Figure 1 by the instruction sequence $inst(store_1) \cdot inst(load_1) := q_{0,1} \xrightarrow{mem[x] \leftarrow 1} q_{1,1} \xrightarrow{r_1 \leftarrow mem[y]} q_{1,2}$.

3.1 Soundness and Completeness

We show that Algorithm 1 is a decision procedure for acyclic programs. From here until (inclusively) Theorem 3 we assume that all programs are acyclic, i.e., their instructions and control states form directed acyclic graphs. Theorem 4 then explains how Algorithm 1 yields a semi-decision procedure for all programs.

We first prove the extension sound and complete (Lemma 1): extending R by sequence $\sigma := \mathcal{O}(R)$ does neither add nor remove TSO-reachable states. Afterwards, Lemma 2 shows that if Algorithm 1 extends R by σ (Line 8) then, in subsequent iterations of the algorithm, no new sequence returned by the oracle is the same as σ (projected back to P). Next, by the first condition of an oracle and using Lemma 2, we establish that Algorithm 1 is a decision procedure for acyclic programs (Theorem 3). Finally, we show that Algorithm 1 can be turned into a semi-decision procedure for all programs using a bounded model checking approach (Theorem 4).

Lemma 1 *Let $DOM \cup REG$ be the addresses and registers of program R and let $\sigma := \mathcal{O}(R)$. Then we have $(pc, val, buf) \in Reach_{TSO}(R)$ if and only if $(pc, val', buf) \in Reach_{TSO}(R \oplus \sigma)$ with $val(a) = val'(a)$ for all $a \in DOM \cup REG$.*

Let t be the thread that differs in R and $R \oplus \sigma$. To prove Lemma 1, one can show that for any prefix α' of $\alpha \in \mathcal{C}_{TSO}(R)$ there is a prefix β' of $\beta \in \mathcal{C}_{TSO}(R \oplus \sigma)$, and vice versa, that maintain the following invariants.

Inv-0 $s_0 \xrightarrow{\alpha'} (pc, val, buf)$ and $s_0 \xrightarrow{\beta'} (pc', val', buf')$.

Inv-1 If pc and pc' differ, they only differ for thread t . If $pc(t) \neq pc'(t)$, then $pc(t) = dst(inst_i)$ and $pc'(t) = \bar{q}_i$ for some $i \in [1..n-1]$.

Inv-2 $val'(a) = val(a)$ for all $a \in DOM \cup REG$.

Inv-3 buf and buf' differ at most for t . If $buf(t) \neq buf'(t)$, then $pc'(t) = \bar{q}_i$ for some $i \in [1..n-1]$ and $buf(t) = (\widehat{ar_{count}}, \widehat{vr_{count}}) \cdots (\widehat{ar_1}, \widehat{vr_1}) \cdot buf'(t)$ where $count$ stores are seen along σ from $src(inst_1)$ to $dst(inst_i)$.

We now show that the oracle never suggests the same sequence σ twice. Since in $R \oplus \sigma$ we introduce new instructions that correspond to instructions in R , we have to map back sequences of instructions I_\oplus in $R \oplus \sigma$ to sequences of instructions I in R . Intuitively, the mapping gives the original instructions from which the sequence was produced. Formally, we define a family of projection functions $h_\sigma: I_\oplus^* \rightarrow I^*$ with $h_\sigma(\varepsilon) := \varepsilon$ and $h_\sigma(w \cdot inst) := h_\sigma(w) \cdot h_\sigma(inst)$. For an instruction $inst \in I_\oplus$, we define $h_\sigma(inst) := inst$ provided $inst \in I$. We set $h_\sigma(inst) := inst_i$ if $inst$ is a first instruction on the path between \bar{q}_{i-1} and \bar{q}_i for some $i \in [1..n]$. In all other cases, we delete the instruction, $h_\sigma(inst) := \varepsilon$. Then, if $R_0 := P$ is the original program, σ_j is the sequence that the oracle returns in iteration $j \in \mathbb{N}$ of the while loop, and w is a sequence of instructions in R_{j+1} , we define $h(w) := h_{\sigma_0}(\dots h_{\sigma_j}(w))$. This latter function maps sequences of instructions in program R_{j+1} back to sequences of instructions in P .

We are ready to state our key lemma. Intuitively, if the oracle in Algorithm 1 returns $\sigma := \mathcal{O}(R)$ and $\sigma' := \mathcal{O}(R \oplus \sigma)$ then, necessarily, $h(\sigma') \neq h(\sigma)$.

Lemma 2 *Let $R_0 := P$ and $R_{i+1} := R_i \oplus \sigma_i$ for $\sigma_i := \mathcal{O}(R_i)$ as in Algorithm 1. If $\sigma_{j+1} \neq \varepsilon$ then $h(\sigma_{j+1}) \neq h(\sigma_i)$ for all $i \leq j$.*

Proof. Assume, to the contrary, that $h(\sigma_{j+1}) = h(\sigma_i)$ for some $i \leq j$ where $\sigma_{j+1} := \mathcal{O}(R_{j+1})$ and $\sigma_i := \mathcal{O}(R_i)$. Let $inst_{\text{first}}$ be the first (store) instruction and $inst_{\text{last}}$ the last (load) instruction of σ_{j+1} . Similarly, let $inst'_{\text{first}}$ and $inst'_{\text{last}}$ be the first and last instructions of σ_i . Since $h(\sigma_{j+1}) = h(\sigma_i)$ it means that $h(inst_{\text{first}}) = h(inst'_{\text{first}})$ and $h(inst_{\text{last}}) = h(inst'_{\text{last}})$.

However, since all control flows of $R_{i+1} := R_i \oplus \sigma_i$ that recreate $h(inst'_{\text{first}})$ before $h(inst'_{\text{last}})$ also place a fence between the two, no other later sequences that the oracle returns have $h(inst'_{\text{first}})$ come before $h(inst'_{\text{last}})$. This in particular means that $\sigma_{j+1} = \mathcal{O}(R_{j+1})$ where $h(inst_{\text{first}})$ comes before $h(inst_{\text{last}})$ does not exist. In conclusion, the initial assumption is false. \square

We can now prove Algorithm 1 sound and complete for acyclic programs (Theorem 3). Lemma 2 and the assumption that the input program is acyclic ensure that if no goal state is found SC-reachable (Line 4), then Algorithm 1 eventually runs out of sequences σ to return (Line 7). If that is the case, $\mathcal{O}(R)$ returns ε in the last iteration of Algorithm 1. By the first oracle condition, we know that the SC- and TSO-reachable states of R are the same. Hence, no goal state is TSO-reachable in R and, by Lemma 1, no goal state is TSO-reachable in the input program P either. Otherwise, a goal state s is SC-reachable by some computation τ in R_j for some $j \in \mathbb{N}$ and, by Lemma 1, there is a TSO computation in P corresponding to τ that reaches s .

Theorem 3 *For acyclic programs, Algorithm 1 terminates. Moreover, it returns true on input P if and only if $\text{Reach}_{\text{TSO}}(P) \cap G \neq \emptyset$.*

Proof. It is immediate that Algorithm 1 always terminates for acyclic programs. On the one hand, the number of instruction sequences that start with a store and end with a load as in the second oracle condition are finite in P . On the

other hand, by Lemma 2, at each iteration the oracle returns a sequence that differs (in P) from the previous ones. These two facts imply termination.

We now prove that $\text{Reach}_{\text{TSO}}(P) \cap G \neq \emptyset$ iff. Algorithm 1 returns *true* on input P . For the easy direction, assume that Algorithm 1 returns *true* on input P . This means that $\text{Reach}_{\text{SC}}(R) \cap G \neq \emptyset$ in the last iteration of the algorithm's loop. Then, by $\text{Reach}_{\text{SC}}(R) \subseteq \text{Reach}_{\text{TSO}}(R)$ and Lemma 1, we know that $\text{Reach}_{\text{SC}}(R) \subseteq \text{Reach}_{\text{TSO}}(P)$. Hence, $\text{Reach}_{\text{TSO}}(P) \cap G \neq \emptyset$.

For the reverse direction, assume that $\text{Reach}_{\text{TSO}}(P) \cap G \neq \emptyset$. Furthermore, let $R_0 := P$ and $R_{i+1} := R_i \oplus \sigma_i$ for $\sigma_i := \mathcal{O}(R_i)$. By the initial termination argument we know there exists $j \in \mathbb{N}$ such that the algorithm terminates with $R = R_j$ in its last loop iteration. That means that either the check in Line 3 of the algorithm succeeds, in which case Algorithm 1 returns *true*, or the check in Line 7 of the algorithm fails, i.e. $\mathcal{O}(R_j) = \epsilon$ and $\text{Reach}_{\text{SC}}(R_j) \cap G = \emptyset$. In the latter case, by the first oracle condition we know that $\text{Reach}_{\text{TSO}}(R_j) \cap G = \emptyset$ and, by Lemma 1, we get $\text{Reach}_{\text{TSO}}(R_j) \subseteq \text{Reach}_{\text{TSO}}(R_0)$. Then, $\text{Reach}_{\text{TSO}}(P) \cap G = \emptyset$ contradicts the above assumption and concludes the proof. \square

To establish that Algorithm 1 is a semi-decision procedure for all programs, one can use an iterative bounded model checking approach. Bounded model checking unrolls the input program P up to a bound $k \in \mathbb{N}$ on the length of computations. Then Algorithm 1 is applied to the resulting programs P_k . If it finds a goal state TSO-reachable in P_k , this state corresponds to a TSO-reachable goal state in P . Otherwise, we increase k and try again. By Theorem 3, we know that Algorithm 1 is a decision procedure for each P_k . This implies that Algorithm 1 together with iterative bounded model checking yields a semi-decision procedure that terminates for all positive instances of TSO reachability. For negative instances of TSO reachability, however, the procedure is guaranteed to terminate only if the input program P is acyclic.

Theorem 4 *We have $G \cap \text{Reach}_{\text{TSO}}(P) \neq \emptyset$ if and only if, for large enough $k \in \mathbb{N}$, Algorithm 1 returns true on input P_k .*

Proof. Assume that $G \cap \text{Reach}_{\text{TSO}}(P) \neq \emptyset$. Then there exist some state $s \in G$ and $\alpha \in \mathcal{C}_{\text{TSO}}(P)$ such that $s_0 \xrightarrow{\alpha} s$. Let k be the length of α and G' be the goal states of $X_{\text{TSO}}(P_k)$. There exists a computation $\beta \in \mathcal{C}_{\text{TSO}}(P_k)$ that mimics α and reaches $s' \in G'$. Hence, $G' \cap \text{Reach}_{\text{TSO}}(P_k) \neq \emptyset$ and, by Theorem 3, Algorithm 1 returns *true* on input P_k .

For the reverse direction, assume that Algorithm 1 returns *true* on input P_k for some $k \in \mathbb{N}$. Let s'_0 be the initial state of $X_{\text{TSO}}(P_k)$ and, as before, G' be the goal states of $X_{\text{TSO}}(P_k)$. By Theorem 3, there exists $s' \in G' \cap \text{Reach}_{\text{TSO}}(P_k)$ and $\beta \in \mathcal{C}_{\text{TSO}}(P_k)$ such that $s'_0 \xrightarrow{\beta} s'$. Since P_k unrolls P up to bound k , there exists a computation $\alpha \in \mathcal{C}_{\text{TSO}}(P)$ that mimics β and reaches $s \in G$. Therefore, $G \cap \text{Reach}_{\text{TSO}}(P) \neq \emptyset$. \square

4 A Robustness-based Oracle

This section argues why robustness yields an oracle. Robustness [7, 10, 13, 26] is a correctness criterion requiring that for each TSO computation of a program there is an SC computation that has the same data and control dependencies. Delays due to store buffering are still allowed, as long as they do not produce dependencies between instructions that SC computations forbid.

Dependencies between events are described in terms of the *happens-before* relation of a computation $\tau \in \mathcal{C}_{\text{TSO}}(P)$. The happens-before relation is a union of the three relations that we define below: $\rightarrow_{hb}(\tau) := \rightarrow_{po} \cup \leftrightarrow \cup \rightarrow_{cf}$.

The *program order relation* \rightarrow_{po} is the order in which threads issue their commands. Formally, it is the union of the program order relations for all threads: $\rightarrow_{po} := \bigcup_{t \in \text{TID}} \rightarrow_{po}^t$. Let τ' be the subsequence of all non-flush events of thread t in τ . Then $\rightarrow_{po}^t := \prec_{\tau'}$.

The *equivalence relation* \leftrightarrow links, in each thread, flush events and their matching store events: $(t, \text{inst}, a) \leftrightarrow (t, \text{flush}, a)$.

The *conflict relation* \rightarrow_{cf} orders accesses to the same address. Assume, on the one hand, that $\tau = \tau_1 \cdot \text{store} \cdot \tau_2 \cdot \text{load} \cdot \tau_3 \cdot \text{flush} \cdot \tau_4$ such that $\text{store} \leftrightarrow \text{flush}$, events store and load access the same address a and come from thread t , and there is no other store event $\text{store}' \in \tau_2$ such that $\text{thread}(\text{store}') = t$ and $\text{addr}(\text{store}') = a$. Then the load event load is an *early read* of the value buffered by the event store and $\text{store} \rightarrow_{cf} \text{load}$.

On the other hand, assume $\tau = \tau_1 \cdot \mathbf{e} \cdot \tau_2 \cdot \mathbf{e}' \cdot \tau_3$ such that \mathbf{e} and \mathbf{e}' are either load or flush events that access the same address a , neither \mathbf{e} nor \mathbf{e}' is an early read, and at least one of \mathbf{e} or \mathbf{e}' is a flush to a . If there is no other flush event $\text{flush} \in \tau_2$ with $\text{addr}(\text{flush}) = a$ then $\mathbf{e} \rightarrow_{cf} \mathbf{e}'$.

Figure 4 depicts the happens-before relation of computation τ_{wit} .

A program P is said to be *robust* against TSO if for each computation $\tau \in \mathcal{C}_{\text{TSO}}(P)$ there exists a computation $\tau' \in \mathcal{C}_{\text{SC}}(P)$ such that $\rightarrow_{hb}(\tau) = \rightarrow_{hb}(\tau')$. If a program P is robust, then it reaches the same set of final states under SC and under TSO:

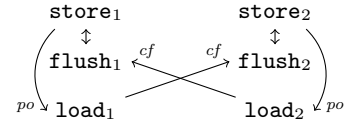


Fig. 4. The relation $\rightarrow_{hb}(\tau_{\text{wit}})$.

Lemma 5 *If P is robust against TSO, then $\text{Reach}_{\text{SC}}(P) = \text{Reach}_{\text{TSO}}(P)$.*

Proof. The \subseteq inclusion holds by $\mathcal{C}_{\text{SC}}(P) \subseteq \mathcal{C}_{\text{TSO}}(P)$. For the reverse, assume that there is a TSO computation $\tau \in \mathcal{C}_{\text{TSO}}(P)$ such that $s_0 \xrightarrow{\tau} s$. Since P is robust, there is an SC computation $\tau' \in \mathcal{C}_{\text{SC}}(P)$ such that $\rightarrow_{hb}(\tau) = \rightarrow_{hb}(\tau')$. Then $\tau' \in \mathcal{C}_{\text{TSO}}(P)$ and, by Lemma 8, $s_0 \xrightarrow{\tau'} s$ so s is SC-reachable. \square

Our robustness-based oracle makes use of the following characterization of robustness from earlier work [10]: a program P is not robust against TSO iff $\mathcal{C}_{\text{TSO}}(P)$ contains a computation, called *witness*, as in Figure 5.

Lemma 6 ([10]) *Program P is robust against TSO if and only if the set of TSO computations $\mathcal{C}_{\text{TSO}}(P)$ contains no witness.*

A witness τ delays stores of only one thread in P . The other threads adhere to the SC semantics. Conditions (W1) – (W4) in Figure 5 describe formally this restrictive behavior. Furthermore, condition (W5) implies that no computation $\tau' \in \mathcal{C}_{SC}(P)$ can satisfy $\rightarrow_{hb}(\tau) = \rightarrow_{hb}(\tau')$.

The computation τ_{wit} is a witness for the program in Figure 1. Indeed, in no SC computation of this program can both loads read the initial values of x and y . Relative to Figure 5, we have $\mathbf{store} = \mathbf{store}_1$, $\mathbf{load} = \mathbf{load}_1$, $\mathbf{flush} = \mathbf{flush}_1$, $\tau_3 = \mathbf{store}_2 \cdot \mathbf{flush}_2 \cdot \mathbf{load}_2$, and $\tau_1 = \tau_2 = \tau_4 = \varepsilon$.

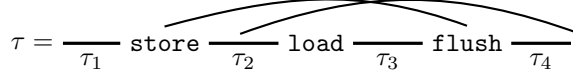


Fig. 5. Witness τ with $\mathbf{store} \leftrightarrow \mathbf{flush}$ and thread $t := \mathit{thread}(\mathbf{store}) = \mathit{thread}(\mathbf{load})$. Witnesses satisfy the following constraints: (W1) Only thread t delays stores. (W2) Event \mathbf{flush} is the first delayed store of t and \mathbf{load} is the last event of t past which \mathbf{flush} is delayed. So τ_2 contains neither flush events nor fences of t . (W3) Sequence τ_3 contains no events of thread t . (W4) Sequence τ_4 consists only of flush events e of thread t . All these events e satisfy $\mathit{addr}(e) \neq \mathit{addr}(\mathbf{load})$. (W5) We require $\mathbf{load} \rightarrow_{hb}^+ e$ for all events e in $\tau_3 \cdot \mathbf{flush}$.

The *robustness-based oracle*, given input P , finds a witness τ as in Figure 5 and returns the sequence of instructions for the events in $\mathbf{store} \cdot \tau_2 \cdot \mathbf{load}$ that belong to thread t . If no witness exists, it returns ε . By Lemmas 5 and 6, this satisfies the oracle conditions from Section 3. Note that, given a robust program and the robustness-based oracle as inputs, Algorithm 1 returns within the first iteration of the while loop.

5 Experiments

We have implemented our lazy TSO reachability algorithm on top of the tool TRENCHER [1]. TRENCHER was initially developed for checking robustness and implements the algorithm for finding witness computations described in [10]. Our implementation reuses that algorithm as a robustness-based oracle. TRENCHER originally used SPIN [17] as back-end SC reachability checker. The current implementation, however, uses a simpler model checker that exploits information about the instruction set for partial-order reduction. Moreover, it avoids having to compile the verifier executables (pan) as is the case for SPIN.

We have implemented Algorithm 1 with the following amendments. First, the extension does not delete the store instruction inst_1 . This ensures the extended program has a (sound) superset of the TSO behaviors of the original program. Second, the extension only adds instructions along $\bar{q}_1, \dots, \bar{q}_n$. The remaining instructions were added to ensure all behaviors of the original program exist in the extended program, once inst_1 is removed. The resulting algorithm is guaranteed to give correct results for cyclic programs. Of course, it cannot be guaranteed to terminate in general. Finally, our implementation explores extensions due to different instruction sequences in parallel, rather than sequentially.

We compare our prototype implementation against two other model checkers that support TSO semantics: MEMORAX [2] (revision 4f94ab6) and CBMC [14] (version 4.7). MEMORAX implements a sound and complete reachability checking procedure by reducing to coverability in a well-structured transition system. CBMC is an SMT-based bounded model checker for C programs. Consequently, it is sound, but not complete: it is complete only up to a given bound on the number of loop iterations in the input program.

5.1 Examples

We tested our tool on a set of examples. Figure 6 summarizes characteristics of the examples taken from the initial TRENCHER tests: number of threads (T), states (St), and transitions (Tr). The first example is a model of the buggy Parker class from Java VM [15]. The next three examples are mutual exclusion protocols implemented via shared variables. These protocols do not guarantee mutual exclusion under TSO. We tested Dekker’s and Peterson’s algorithms for two threads, and Lamport’s fast mutex [22] for three threads. The last three tests from Figure 6 give statistics concerning reachability in robust test cases for the lock-free stack, and for the MCS and CLH locking algorithms from [16].

We also performed three parametrized tests. First, we varied the number of threads in Lamport’s fast mutex [22] (see left-hand-side of Figure 7). The modified Dekker in Figure 8 is inspired by the examples of the fence-insertion tool MUSKETEEER [4] and adds an “ N -branching diamond” (see right-hand-side of Figure 8) to both program threads. Lastly, the program in Figure 9 places stores to address x on a length N loop in thread t_1 : since t_1 expects to load the initial y value while t_2 expects to load 1 and then 0 from x , an execution that reaches the goal state goes through the length N loop twice.

5.2 Evaluation

We ran all tests on a QEMU @ 2.67GHz virtual machine (16 cores) with 8GB RAM running GNU/Linux. The table in Figure 6 summarizes the results of the TRENCHER benchmark tests. RQ is the number of SC reachability queries raised by TRENCHER. The columns CPU and Real give the total CPU time and the wall-clock time for performing a test.

The first graph in Figure 10 depicts the running times of the three tools on the non-robust examples from Figure 6. For CBMC, we used the versions of the mutual exclusion algorithms that its authors provide. For MEMORAX, we hand-wrote *.rmm files for the first 4 test programs. We did not perform a comparison for robust programs: if SC reachability returns false on an input

#	Program	T	St	Tr	RQ	CPU	Real
1	Parker (non-rob)	2	11	10	4	8	5
2	Peterson (non-rob)	2	14	18	12	21	13
3	Dekker (non-rob)	2	24	30	30	171	70
4	Lamport (non-rob)	3	33	36	27	1839	694
5	MCS Lock	4	52	50	30	127	61
6	CLH Lock	3	43	41	70	10	7
7	Lock-Free Stack	4	46	50	14	9	7

Fig. 6. TRENCHER benchmarking results. The tests are available online [1]. Times here are in milliseconds.

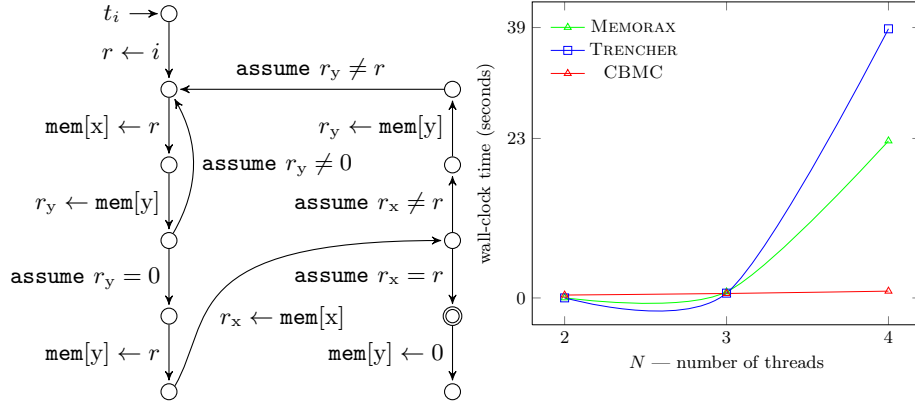


Fig. 7. The i -th Lamport mutex thread (left) and running times for N threads (right).

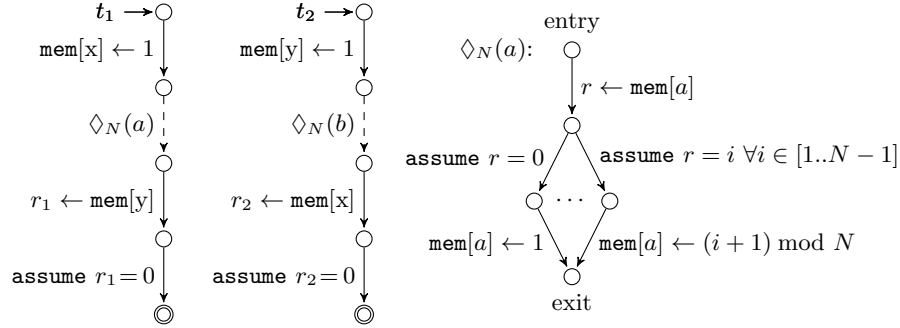


Fig. 8. Dekker's algorithm modified so that an “ N -branching diamond” over distinct addresses $a, b \notin \{x, y\}$ is placed between the accesses to x and y . A final goal state is TSO-reachable if the first store is delayed past the last load in either t_1 or t_2 .

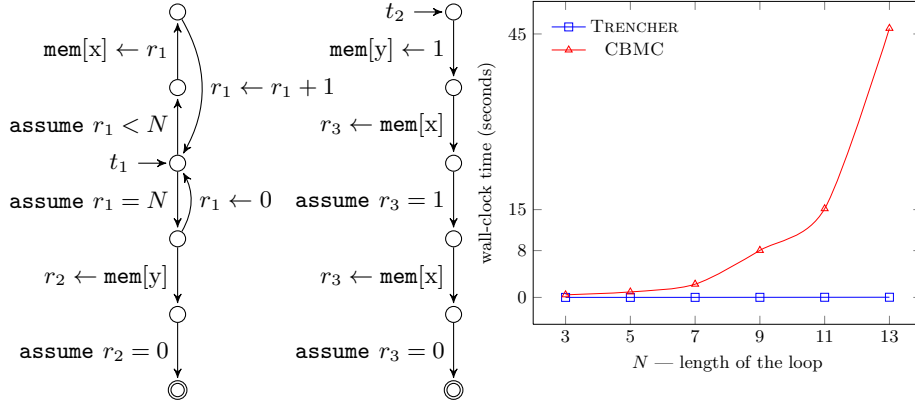


Fig. 9. A final goal state is TSO-reachable if t_1 goes through the (length N) loop two times: once to satisfy $\text{assume } r_3 = 1$ and the second time to satisfy $\text{assume } r_3 = 0$.

program, our implementation decides mutual exclusion as fast as TRENCHER is

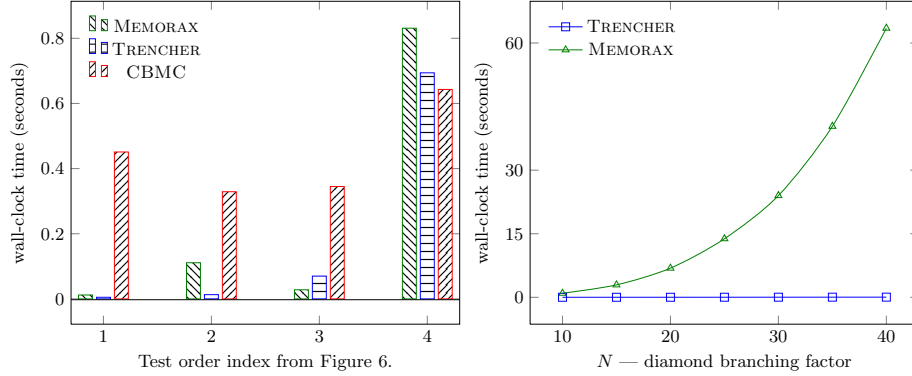


Fig. 10. Running times for the non-robust tests in Figure 6 (left) and Figure 8 (right).

able to determine robustness. Moreover, CBMC implements strictly an under-approximative method where the number of loop iterations is bounded. Our robust tests, however, contain unbounded loops.

The high load needed to verify Lamport’s mutex — in comparison with the other Figure 6 tests — is justified by the correlation between the program’s data domain size and its number of threads. For a larger number of threads, the right-hand-side graph in Figure 7 shows that CBMC is fastest. This is the case since, actually, the smallest unwind bound suffices for CBMC to conclude reachability. For MEMORAX and TRENCHER the system runs out of memory when $N = 5$. This underlines once again just how troublesome the state-space explosion is for TSO reachability. Although it is not easily noticeable in the picture, MEMORAX’s exponential scaling is better than TRENCHER’s: although TRENCHER is slightly faster than MEMORAX for $N \in \{2, 3\}$, MEMORAX clearly outperforms TRENCHER when $N = 4$.

The graph in Figures 9 show that, for the second parameterized test, our prototype is faster than CBMC. Indeed, with increasing N , an ever larger number of constraints need to be generated by CBMC. For TRENCHER, regardless of the value of N , it takes three SC reachability queries to conclude TSO reachability.

The second graph in Figure 10 shows that, for the programs described by Figure 8, our prototype is faster than MEMORAX. It seems MEMORAX cannot cope well with the branching factor that the parameter N introduces.

To better understand the difficulty of the latter two parametric tests, we present the exponential scaling behaviors of TRENCHER in Figure 11.

5.3 Discussion

Because we find several witnesses in parallel, throughout the experiments our implementation required up to 2 iterations of the loop in Algorithm 1. In the case of robust programs, one iteration is always sufficient. This suggests that robustness violations are really the critical behaviors leading to TSO reachability.

The experiments indicate that, at least for some programs with a high branching factor, our implementation is faster than MEMORAX if a useful witness can be

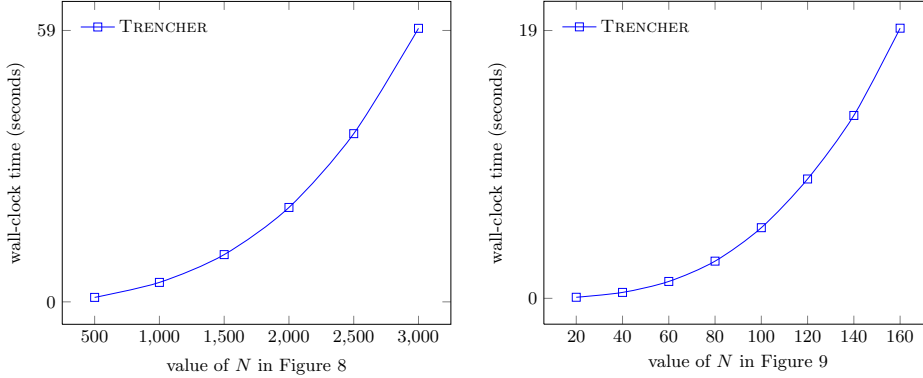


Fig. 11. Additional TRENCHER results for the programs in Figures 8 and 9. MEMORAX takes already 1 minute and 24 seconds for the program in Figure 8 and $N = 50$, while CBMC takes 8 minutes and 35 seconds for the program in Figure 9 and $N = 20$.

found within a small number of iterations of Algorithm 1. Similarly, our prototype is better than CBMC for programs which require a high unwinding bound to make visible TSO behavior reaching a goal state. Although the two programs by which we show this are rather artificial, we expect such characteristics to occur in actual code. Hence, our approach seems to be strong on an orthogonal set of programs. In a portfolio model checker, it could be used as a promising alternative to the existing techniques.

To evaluate the practicality of our method, more experiments are needed. In particular, we hope to be able to substantiate the above conjecture for concrete programs with behavior like that depicted in Figures 8 and 9. Unfortunately, there seems to be no clear way of translating (compiled) C programs into our simplified assembly syntax without substantial abstraction. To handle C code, an alternative would be to reimplement our method within CBMC. But this would force us to determine a-priori a good-enough unwinding bound. Moreover, we could no longer conclude safety of robust programs with unbounded loops.

Acknowledgements The third author was granted by the Competence Center High Performance Computing and Visualization (CC-HPC) of the Fraunhofer Institute for Industrial Mathematics (ITWM). The work was partially supported by the PROCOPE project ROIS: *Robustness under Realistic Instruction Sets* and by the DFG project R2M2: *Robustness against Relaxed Memory Models*.

References

1. Trencher tool. <http://concurrency.informatik.uni-kl.de/trencher.html>.
2. P. A. Abdulla, M. F. Atig, Y.-F. Chen, C. Leonardsson, and A. Rezne. Counter-Example Guided Fence Insertion under TSO. In *TACAS*, volume 7214 of *LNCS*, pages 204–219. Springer, 2012.
3. J. Alglave. *A Shared Memory Poetics*. PhD thesis, University Paris 7, 2010.

4. J. Alglave, D. Kroening, V. Nimal, and D. Poetzl. Don't Sit on the Fence - A Static Analysis Approach to Automatic Fence Insertion. In *CAV*, volume 8559 of *LNCS*, pages 508–524. Springer, 2014.
5. J. Alglave, D. Kroening, V. Nimal, and M. Tautschnig. Software Verification for Weak Memory via Program Transformation. In *ESOP*, volume 7792 of *LNCS*, pages 512–532. Springer, 2013.
6. J. Alglave, D. Kroening, and M. Tautschnig. Partial Orders for Efficient BMC of Concurrent Software. *CoRR*, abs/1301.1629, 2013.
7. J. Alglave and L. Maranget. Stability in Weak Memory Models. In *CAV*, volume 6806 of *LNCS*, pages 50–66. Springer, 2011.
8. M. F. Atig, A. Bouajjani, S. Burckhardt, and M. Musuvathi. On the Verification Problem for Weak Memory Models. In *POPL*, pages 7–18. ACM, 2010.
9. M. F. Atig, A. Bouajjani, and G. Parlato. Getting Rid of Store-Buffers in TSO Analysis. In *CAV*, volume 6806 of *LNCS*, pages 99–115. Springer, 2011.
10. A. Bouajjani, E. Derevenetc, and R. Meyer. Checking and Enforcing Robustness against TSO. In *ESOP*, volume 7792 of *LNCS*, pages 533–553. Springer, 2013.
11. A. Bouajjani, R. Meyer, and E. Möhlmann. Deciding Robustness against Total Store Ordering. In *ICALP*, volume 6756 of *LNCS*, pages 428–440. Springer, 2011.
12. S. Burckhardt and M. Musuvathi. Effective Program Verification for Relaxed Memory Models. In *CAV*, volume 5123 of *LNCS*, pages 107–120. Springer, 2008.
13. J. Burnim, C. Stergiou, and K. Sen. Sound and Complete Monitoring of Sequential Consistency for Relaxed Memory Models. In *TACAS*, volume 6605 of *LNCS*, pages 11–25. Springer, 2011.
14. E. Clarke, D. Kroening, and F. Lerda. A Tool for Checking ANSI-C Programs. In *TACAS*, volume 2988 of *LNCS*, pages 168–176. Springer, 2004.
15. D. Dice. A race in LockSupport park() arising from Weak Memory Models. https://blogs.oracle.com/dave/entry/a_race_in_locksupport_park.
16. M. Herlihy and N. Shavit. *The Art of Multiprocessor Programming*. MKP, 2008.
17. G. J. Holzmann. The Model Checker SPIN. *IEEE Tr. Sof. Eng.*, 23:279–295, 1997.
18. D. Kozen. Lower Bounds for Natural Proof Systems. In *FOCS*, pages 254–266. IEEE Computer Society, 1977.
19. M. Kuperstein, M. Vechev, and E. Yahav. Partial-Coherence Abstractions for Relaxed Memory Models. In *PLDI*, pages 187 – 198. ACM, 2011.
20. M. Kuperstein, M. T. Vechev, and E. Yahav. Automatic Inference of Memory Fences. *ACM SIGACT News*, 43(2):108–123, 2012.
21. L. Lamport. How to Make a Multiprocessor Computer that Correctly Executes Multiprocess Programs. *IEEE Tr. on Com.*, 28(9):690–691, 1979.
22. L. Lamport. A Fast Mutual Exclusion Algorithm. *ACM Tr. Com. Sys.*, 5(1), 1987.
23. A. Linden and P. Wolper. An Automata-Based Symbolic Approach for Verifying Programs on Relaxed Memory Models. In *SPIN*, volume 6349 of *LNCS*, pages 212–226. Springer, 2010.
24. A. Linden and P. Wolper. A Verification-based Approach to Memory Fence Insertion in Relaxed Memory Systems. In *MCS*, volume 6823 of *LNCS*, pages 144–160. Springer, 2011.
25. S. Owens, S. Sarkar, and P. Sewell. A Better x86 Memory Model: x86-TSO (extended version). Technical Report CL-TR-745, University of Cambridge, 2009.
26. D. Shasha and M. Snir. Efficient and Correct Execution of Parallel Programs that Share Memory. *ACM Tr. on Prog. Lang. and Sys.*, 10(2):282–312, 1988.

A A Simple Safe Program

The program from Figure 12 is safe since no goal state is TSO-reachable: the initial control states will never be left since the conditionals will never succeed. However, the algorithm that we describe for Theorem 4 does not terminate for this example. Although every P_k that unrolls the program in Figure 12 up to $k \in \mathbb{N}$ is found safe, the algorithm only stops if a TSO-reachable state is found or if $\mathcal{O}(R) = \epsilon$, which is never the case.

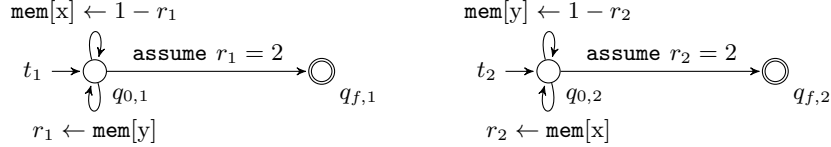


Fig. 12. A safe program for which Algorithm 1 (as in Theorem 4) does not terminate.

The underlying reason why always $\mathcal{O}(R) \neq \epsilon$ is that there are infinitely many sequences $inst_{store}^m \cdot inst_{load}$, where $inst_{store} = (q_{0,1}, \text{mem}[x] \leftarrow 1 - r_1, q_{0,1})$, $inst_{load} = (q_{0,1}, r_1 \leftarrow \text{mem}[y], q_{0,1})$, and $m \in \mathbb{N}$.

B Proofs missing in Subsection 3.1

Prior to proving Lemma 1 we do a bit of preparation. We rely on computations that delay flush events locally the least. Lemma 7 explains what this means.

Lemma 7 *Let $\alpha \in \mathcal{C}_{TSO}(R)$ and $t \in TID$. There exists $\ddot{\alpha} \in \mathcal{C}_{TSO}(R)$ such that $\rightarrow_{hb}(\alpha) = \rightarrow_{hb}(\ddot{\alpha})$ and, for all events $e_{store} \leftrightarrow e_{flush}$ within thread t , if $\ddot{\alpha} \downarrow t := \alpha_{prefix} \cdot e_{store} \cdot \alpha' \cdot e_{flush} \cdot \alpha_{suffix}$ then either*

- (1) $\alpha' := \beta \cdot e_{load} \cdot \beta'$ and all events $e \in \beta'$ are flushes,
- or (2) all events $e \in \alpha'$ are local assignments or conditionals

Proof. Intuitively, the theorem states that flush events of thread t delayed past same-thread local events, may be delayed less without changing the happens-before relation of the computation. Local events are assignments, conditionals, and store events in the same thread.

Let $\alpha := \alpha_1 \cdot e_{store} \cdot \alpha_2 \cdot e \cdot \alpha_3 \cdot e_{flush} \cdot \alpha_4$ such that $e_{store} \leftrightarrow e_{flush}$ are events of thread t , e is a local event in t and $thread(e') \neq t$ for all events $e' \in \alpha_3$.

We denote by $\alpha_0 := \alpha_1 \cdot e_{store} \cdot \alpha_2 \cdot \alpha_3 \cdot e_{flush} \cdot e \cdot \alpha_4$ the TSO computation that first performs the flush e_{flush} and then the event e . Notice that since α_3 contains no events e' with $thread(e') = t$, feasibility of computation α_0 is ensured and $\rightarrow_{hb}(\alpha) = \rightarrow_{hb}(\alpha_0)$ holds.

Starting with the last flush event in α , we use the above reordering of events e to locally delay flush events less. In the end we obtain computation $\ddot{\alpha}$ in which no flush event of thread t can be locally delayed less. \square

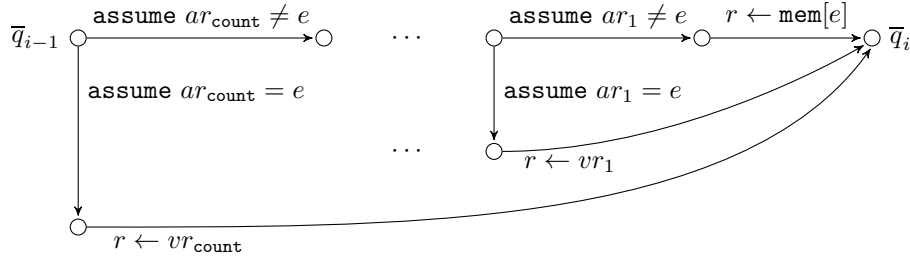
Furthermore, in order to reference instructions of $R \oplus \sigma$ that the extension adds we give an alternative description for some of the transition sequences in

the main text. Recall that variable `count` keeps track of the number of store instructions processed along σ .

If $\text{cmd}(\text{inst}_i) = \text{mem}[e] \leftarrow e'$, we said `count` is incremented and instructions that remember the value and address written in ar_{count} and vr_{count} are added.

$$\bar{q}_{i-1} \xrightarrow{\text{ar}_{\text{count}} \leftarrow e} \circ \xrightarrow{\text{vr}_{\text{count}} \leftarrow e'} \circ \bar{q}_i \quad (1)$$

If $\text{cmd}(\text{inst}_i) = r \leftarrow \text{mem}[e]$ we said instructions are added that load from memory only when a load from the simulated buffer is not possible. More precisely, if some $j \in [1, \text{count}]$ such that $\text{ar}_j = e$ is found, r is assigned the value of vr_j . Otherwise, the register r receives its value from the address \hat{e} .



Alternatively, assuming $\bar{q}_{\text{check},i,\text{count}} := \bar{q}_{i-1}$, this can be stated as adding

$$\{(\bar{q}_{\text{check},i,\text{count}}, \text{assume } \text{ar}_{\text{count}} = e, \bar{q}_{\text{buf},i,\text{count}})\} \quad (2)$$

$$\uplus \{(\bar{q}_{\text{check},i,\text{count}}, \text{assume } \text{ar}_{\text{count}} \neq e, \bar{q}_{\text{check},i,\text{count}-1})\} \quad (3)$$

$$\uplus \{(\bar{q}_{\text{buf},i,\text{count}}, r \leftarrow \text{vr}_{\text{count}}, \bar{q}_i)\} \quad (4)$$

\vdots

$$\uplus \{(\bar{q}_{\text{check},i,1}, \text{assume } \text{ar}_1 = e, \bar{q}_{\text{buf},i,1})\} \quad (5)$$

$$\uplus \{(\bar{q}_{\text{check},i,1}, \text{assume } \text{ar}_1 \neq e, \bar{q}_{\text{mem},i})\} \quad (6)$$

$$\uplus \{(\bar{q}_{\text{buf},i,1}, r \leftarrow \text{vr}_1, \bar{q}_i)\} \quad (7)$$

$$\uplus \{(\bar{q}_{\text{mem},i}, r \leftarrow \text{mem}[e], \bar{q}_i)\} \quad (8)$$

We said that out of control state \bar{q}_n we create a sequence of stores to flush the contents of the auxiliary registers and return to the code of the original thread.

$$\bar{q}_n \xrightarrow{\text{mem}[\text{ar}_1] \leftarrow \text{vr}_1} \circ \dots \circ \xrightarrow{\text{mem}[\text{ar}_{\text{max}}] \leftarrow \text{vr}_{\text{max}}} \circ \text{dst}(\text{inst}_n)$$

Alternatively, we could have stated it as adding

$$\{(\bar{q}_n, \text{mem}[\text{ar}_1] \leftarrow \text{vr}_1, \bar{q}_{\text{flush},1})\} \quad (9)$$

\vdots

$$\uplus \{(\bar{q}_{\text{flush},\text{max}-1}, \text{mem}[\text{ar}_{\text{max}}] \leftarrow \text{vr}_{\text{max}}, \text{dst}(\text{inst}_n))\} \quad (10)$$

Furthermore, for all instructions $inst \in I_t$ with $src(inst) = src(inst_i)$ for some $i \in [1..n]$ and for which $inst \neq inst_i$ we added instructions that flush the stores buffered in the auxiliary registers and return to $dst(inst)$.

$$\bar{q}_i \xrightarrow{\text{mem}[ar_1] \leftarrow vr_1} \circ \dots \circ \xrightarrow{\text{mem}[ar_{count}] \leftarrow vr_{count}} \circ \xrightarrow{cmd(inst)} \circ dst(inst)$$

Alternatively, we could have stated it as adding

$$\{(\bar{q}_i, \text{mem}[ar_1] \leftarrow vr_1, \bar{q}_{\text{next},i,1})\} \quad (11)$$

\vdots

$$\uplus \{(\bar{q}_{\text{next},i,\text{count}-1}, \text{mem}[ar_{count}] \leftarrow vr_{count}, \bar{q}_{\text{next},i,\text{count}})\} \quad (12)$$

$$\uplus \{(\bar{q}_{\text{next},i,\text{count}}, cmd(inst), dst(inst))\} \quad (13)$$

Finally, for all load instructions $inst_i$, where $i < n$, as well as out of \bar{q}_1 we added instructions that flush and fence the pair (ar_1, vr_1) , make the remaining buffered stores in the auxiliary registers visible, and return to q . Here $q := src(inst_i)$ in the load case and $q := dst(inst_1)$ otherwise.

$$\bar{q}_i \xrightarrow{\text{mem}[ar_1] \leftarrow vr_1} \circ \xrightarrow{\text{mfence}} \circ \dots \circ \xrightarrow{\text{mem}[ar_{count}] \leftarrow vr_{count}} \circ q$$

Alternatively, we could have stated it as adding

$$\{(\bar{q}_i, \text{mem}[ar_1] \leftarrow vr_1, \bar{q}_{\text{fence},i})\} \quad (14)$$

$$\uplus \{(\bar{q}_{\text{fence},i}, \text{mfence}, \bar{q}_{\text{orig},i,2})\} \quad (15)$$

$$\uplus \{(\bar{q}_{\text{orig},i,2}, \text{mem}[ar_2] \leftarrow vr_2, \bar{q}_{\text{orig},i,3})\} \quad (16)$$

\vdots

$$\uplus \{(\bar{q}_{\text{orig},i,\text{count}}, \text{mem}[ar_{count}] \leftarrow vr_{count}, q)\} \quad (17)$$

We can now turn to the actual proof of Lemma 1.

Proof (of # 1). Assume t is the thread of $\sigma := inst_1 \dots inst_n$, $X_{\text{TSO}}(R \oplus \sigma) := (E_{\oplus}, S_{\oplus}, \Delta_{\text{TSO}}, s_{\oplus}, F_{\oplus})$, I and Q are the instructions and states of R , DOM and REG are registers and addresses used by R , and I_{\oplus} are the instructions I'_t of $R \oplus \sigma$ as described in Section 3.

A direct result of Lemmas 7 and 8 is that TSO computations of R that delay flushes of t locally the least reach all the states in the set $Reach_{\text{TSO}}(R)$. Assume $\alpha \in \mathcal{C}_{\text{TSO}}(R)$ is a computation where flushes of t are delayed locally the least as Lemma 7 describes and let $s_0, \dots, s_m \in S_{\text{TSO}}$ for some $m \in \mathbb{N}$ be all the states along the transition sequence $s_0 \xrightarrow{\alpha} s$, i.e., $s_0 := s_0$ and $s_m := s$. Also, for all $k \in [0, m]$, let α_k denote prefixes of α with $s_0 \xrightarrow{\alpha_k} s_k$.

We prove by induction over state indexes $k \in [0, m]$ that there exist prefixes β_k of $\beta \in \mathcal{C}_{\text{TSO}}(R \oplus \sigma)$ and states $s'_0, \dots, s'_m \in S_{\oplus}$ along $s_{\oplus} \xrightarrow{\beta} s' \in \Delta_{\text{TSO}}^*$ with $s'_0 := s_{\oplus}$ and $s'_m := s'$ such that the following invariants hold:

Inv-0 $s_0 \xrightarrow{\alpha'} (\text{pc}, \text{val}, \text{buf})$ and $s_{\oplus} \xrightarrow{\beta'} (\text{pc}', \text{val}', \text{buf}')$.

Inv-1 If pc and pc' differ then they only differ for thread t . Moreover, if $\text{pc}(t) \neq \text{pc}'(t)$ then $\text{pc}(t) = \text{dst}(\text{inst}_i)$ and $\text{pc}'(t) = \bar{q}_i$ for some $i \in [1..n-1]$.

Inv-2 $\text{val}'(a) = \text{val}(a)$ for all $a \in \text{DOM} \cup \text{REG}$.

Inv-3 buf and buf' differ at most for t . Furthermore, if $\text{buf}(t) \neq \text{buf}'(t)$ then $\text{pc}'(t) = \bar{q}_i$ for some $i \in [1..n-1]$ and $\text{buf}(t) = (\widehat{ar_{\text{count}}}, \widehat{vr_{\text{count}}}) \cdot \dots \cdot (\widehat{ar_1}, \widehat{vr_1}) \cdot \text{buf}'(t)$ where count stores are seen along σ from $\text{src}(\text{inst}_1)$ to $\text{dst}(\text{inst}_i)$.

For the induction **base case** $k = 0$, $\alpha_0 = \epsilon$, $s_0 = s_0$, $\text{pc} = \text{pc}_0$, $\text{val} = \text{val}_0$, and $\text{buf} = \text{buf}_0$. Then, for $\beta_0 := \epsilon$ and $s'_0 = s_{\oplus}$, invariants **Inv-0...3** hold.

For the **induction step case**, assume that invariants **Inv-0...3** hold for $k < m$ and that $s_k \xrightarrow{\mathbf{e}} s_{k+1} := (\text{pc}_+, \text{val}_+, \text{buf}_+)$ for some $\mathbf{e} \in E$. We use a case distinction over possible events \mathbf{e} to define β_{k+1} such that $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1} := (\text{pc}'_+, \text{val}'_+, \text{buf}'_+)$ and invariants **Inv-0...3** hold for $k+1$.

If $\text{thread}(\mathbf{e}) := t' \neq t$ it means $\text{inst}(\mathbf{e}) \in I_{\oplus}$ is enabled in $\text{pc}'(t')$, so there exist $\mathbf{e}' \in E_{\oplus}$ and $s'_{k+1} \in S_{\oplus}$ such that $\text{inst}(\mathbf{e}') := \text{inst}(\mathbf{e})$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ in $X_{\text{TSO}}(R \oplus \sigma)$. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that, by the Δ_{TSO} semantics (Figure 2) and under the assumption that invariants **Inv-0...3** hold for k , invariants **Inv-0...3** also hold for $k+1$.

If $\text{thread}(\mathbf{e}) = t$ we make the following case distinction over \mathbf{e} and $\text{pc}'(t)$.

[1] “ \mathbf{e} is a flush event.” This first case deals with the possibility that a store operation is flushed. Depending on whether $\text{buf}'(t) \neq \epsilon$, we either flush the oldest address-value pair of $\text{buf}'(t)$ or the first address-value auxiliary registers pair. By Lemma 7, the later case can only happen when $\text{pc}'(t) = \bar{q}_i$ for some $i \in [2..n-1]$ and inst_i performs a load or $i = 1$.

If $\text{buf}'(t) \neq \epsilon$ we flush the oldest write access buffered. Namely, let $\mathbf{e}_{\text{flush}} \in E_{\oplus}$ and $s'_{k+1} \in S_{\oplus}$ such that, according to rule (WM), $(s'_k, \mathbf{e}_{\text{flush}}, s'_{k+1}) \in \Delta_{\text{TSO}}$. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}_{\text{flush}}$ and invariants **Inv-0...3** hold for $k+1$ since

(0) **Inv-0,3** hold for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, implying **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{pc}(t)$, and $\text{pc}'_+(t) = \text{pc}'(t)$, so **Inv-1** holds for $k+1$.

(2) **Inv-2,3** hold for k , so events \mathbf{e} and $\mathbf{e}_{\text{flush}}$ update the same address by a same value and **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k and events \mathbf{e} and $\mathbf{e}_{\text{flush}}$ remove one address-value pair from both $\text{buf}(t)$ and $\text{buf}'(t)$, so **Inv-3** holds for $k+1$.

Otherwise, $\text{buf}'(t) = \epsilon$ and count stores are encountered from $\text{src}(\text{inst}_1)$ to $\text{pc}'(t) = \bar{q}_i$ for some $i \in [1..n-1]$. Then $\text{buf}(t) = (\widehat{ar_{\text{count}}}, \widehat{vr_{\text{count}}}) \cdot \dots \cdot (\widehat{ar_1}, \widehat{vr_1})$ and, by Lemma 7, we know inst_i is either the first store inst_1 of σ or a load. Either way, let $\mathbf{e}_1, \dots, \mathbf{e}_{\text{count}}, \mathbf{e}_{\text{flush}}, \mathbf{e}_{\text{fence}} \in E_{\oplus}$ match equations (14–17) in the extension and $s'_{k+1} \in S_{\oplus}$ such that events \mathbf{e}_j are, for all $j \in [1..\text{count}]$, the buffering events for the stores (14,16–17), $\mathbf{e}_{\text{flush}}$ is the flush event for the store (14), $\mathbf{e}_{\text{fence}}$ is the event for the fence (15), and $s'_k \xrightarrow{\mathbf{e}_1 \cdot \mathbf{e}_{\text{flush}} \cdot \mathbf{e}_{\text{fence}} \cdot \mathbf{e}_2 \cdot \dots \cdot \mathbf{e}_{\text{count}}} s'_{k+1} \in \Delta_{\text{TSO}}^*$ according to rules (ST, MEM, F) in Figure 2. We then define $\beta_{k+1} :=$

$\beta_k \cdot \mathbf{e}_1 \cdot \mathbf{e}_{\text{flush}} \cdot \mathbf{e}_{\text{fence}} \cdot \mathbf{e}_2 \cdot \dots \cdot \mathbf{e}_{\text{count}} \cdot \mathbf{e}$ and find that invariants **Inv-0...3** hold for $k+1$ since

(0) **Inv-0,3** hold for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k and $\text{pc}_+(t) = q = \text{pc}'_+(t)$, where $q := \text{src}(\text{inst}_i)$ if inst_i is a load and $q := \text{dst}(\text{inst}_1)$ otherwise, so **Inv-1** holds for $k+1$.

(2) **Inv-2,3** hold for k , events \mathbf{e} and $\mathbf{e}_{\text{flush}}$ update the same address by the same value and, since the other events do not update any address, **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k , and events $\mathbf{e}_2, \dots, \mathbf{e}_{\text{count}}$ place the corresponding address-value pairs that match $\text{buf}_+(t)$ into $\text{buf}'_+(t)$, so **Inv-3** holds for $k+1$.

[2] “ \mathbf{e} is not a flush event, $\text{pc}'(t) = \bar{q}_i$ for $i \in [1..n-1]$, $\text{inst}(\mathbf{e}) \neq \text{inst}_{i+1}$.”

Event \mathbf{e} corresponds to an instruction that does not follow σ . Then, events for instructions (11–13) place the auxiliary address-value pairs into $\text{buf}'_+(t)$ and then perform $\text{cmd}(\text{inst}(\mathbf{e}))$. Let $\mathbf{e}_1, \dots, \mathbf{e}_{\text{count}}, \mathbf{e}' \in E_{\oplus}$ and $s'_{k+1} \in S_{\oplus}$ such that \mathbf{e}_j are, for all $j \in [1..\text{count}]$, the buffering events for stores (11–12), \mathbf{e}' is the event for instruction (13), and $s'_k \xrightarrow{\mathbf{e}_1 \dots \mathbf{e}_{\text{count}} \cdot \mathbf{e}'} s'_{k+1} \in \Delta_{\text{TSO}}^*$, according to the Figure 2 rules. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}_1 \cdot \dots \cdot \mathbf{e}_{\text{count}} \cdot \mathbf{e}'$ and find that invariants **Inv-0...3** hold for $k+1$ since

(0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k and $\text{pc}_+(t) = \text{dst}(\text{inst}(\mathbf{e})) = \text{pc}'_+(t)$, so **Inv-1** holds for $k+1$.

(2) **Inv-2** holds for k and the events \mathbf{e} and \mathbf{e}' update at most one REG register by the same value, so **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k , the buffering store events $\mathbf{e}_1, \dots, \mathbf{e}_{\text{count}}$ make the address-value pairs of the auxiliary registers explicit in $\text{buf}'_+(t)$, and if events \mathbf{e} and \mathbf{e}' are buffering events for stores then they add the same address-value pair, so **Inv-3** holds for $k+1$.

[3] “ $\text{inst}(\mathbf{e})$ performs a store and **[2]** fails.” We analyze the following sub-cases depending on the value of $\text{pc}'(t)$.

[3a] “ $\text{pc}'(t) = \bar{q}_{i-1}$ for some $i \in [1..n-1]$.” Since **[2]** does not hold, $\text{inst}(\mathbf{e}) = \text{inst}_i$ and auxiliary registers track the store inst_i . Let $\mathbf{e}_a, \mathbf{e}_v \in E_{\oplus}$ be events for the instructions in (1) and $s'_{k+1} \in S_{\oplus}$ such that $s'_k \xrightarrow{\mathbf{e}_a \cdot \mathbf{e}_v} s'_{k+1} \in \Delta_{\text{TSO}}^*$ according to the Δ_{TSO} rule for local assignments. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}_a \cdot \mathbf{e}_v$ and find that invariants **Inv-0...3** hold for $k+1$ since

(0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{dst}(\text{inst}_i)$, and $\text{pc}'_+(t) = \bar{q}_i$, so **Inv-1** holds for $k+1$.

(2) **Inv-2** holds for k and no memory changes occurred outside of auxiliary registers, so **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k and $(\widehat{ar_{\text{count}}}, \widehat{vr_{\text{count}}})$ matches the address-value pair added by \mathbf{e} to $\text{buf}_+(t)$, so **Inv-3** holds for $k+1$.

[3b] “ $\text{pc}'(t) = \text{pc}(t) \neq \text{src}(\text{inst}_1)$.” This case is similar to the one when $\text{thread}(\mathbf{e}) \neq t$ since $\text{inst}(\mathbf{e}) \in I_\oplus$. Then there exist $\mathbf{e}' \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that $\text{inst}(\mathbf{e}') = \text{inst}(\mathbf{e})$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ in $X_{\text{TSO}}(R \oplus \sigma)$. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that, by the Δ_{TSO} semantics (Figure 2), invariants **Inv-0...3** continue to hold for $k+1$.

[4] “ $\text{inst}(\mathbf{e})$ performs a load and **[2]** fails.” We analyze the following subcases depending on the value of $\text{pc}'(t)$.

[4a] “ $\text{pc}'(t) = \bar{q}_{i-1}$ for some $i \in [1..n-1]$.” Since **[2]** does not hold, $\text{inst}(\mathbf{e}) = \text{inst}_i$ and we use (4-7,8) to load from e only when no register ar_j matches e for any $j \in [1..\text{count}]$.

If there exists a largest $j \in [1..\text{count}]$ such that $ar_j = e$ then r will take its value from the auxiliary register vr_j . Let $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_j, \mathbf{e}_{\text{assign}} \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that \mathbf{e}_k are, for all $k \in [j+1..\text{count}]$, the events for negative conditional checks (3,6), \mathbf{e}_j is the event for the earliest positive conditional check (2,5), $\mathbf{e}_{\text{assign}}$ is the event for an instruction (4,7), and $s'_k \xrightarrow{\mathbf{e}_{\text{count}} \dots \mathbf{e}_j \cdot \mathbf{e}_{\text{assign}}} s'_{k+1} \in \Delta_{\text{TSO}}^*$ according to the rules for conditionals and local assignments in Δ_{TSO} . We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}_{\text{count}} \dots \mathbf{e}_j \cdot \mathbf{e}_{\text{assign}}$ and find that the invariants **Inv-0...3** hold for $k+1$ since

(0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{dst}(\text{inst}_i)$, and $\text{pc}'_+(t) = \bar{q}_i$, so **Inv-1** holds for $k+1$.

(2) **Inv-2** holds for k , both \mathbf{e} and $\mathbf{e}_{\text{assign}}$ update r by the same value, and no other event $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_j$ changes any address, so **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k and no event alters buffer contents, so **Inv-3** holds for $k+1$.

Otherwise, $ar_j \neq e$ holds for all $j \in [1..\text{count}]$ and the register r will take its value from the address indicated by e . Namely, let $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_1, \mathbf{e}_{\text{load}} \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that \mathbf{e}_k are, for all $k \in [1..\text{count}]$, the events for negative conditional checks (3,6), \mathbf{e}_{load} is the event for instruction (8), and $s'_k \xrightarrow{\mathbf{e}_{\text{count}} \dots \mathbf{e}_1 \cdot \mathbf{e}_{\text{load}}} s'_{k+1} \in \Delta_{\text{TSO}}^*$ according to the rule for conditionals in Δ_{TSO} and (LB/LM). We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}_{\text{count}} \dots \mathbf{e}_1 \cdot \mathbf{e}_{\text{load}}$ and find that invariants **Inv-0...3** hold for $k+1$:

(0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.

(1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{dst}(\text{inst}_i)$, and $\text{pc}'_+(t) = \bar{q}_i$, so **Inv-1** holds for $k+1$.

(2) **Inv-2** holds for k , both \mathbf{e} and \mathbf{e}_{load} update r by the same value, and no other event $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_1$ changes any address, so **Inv-2** holds for $k+1$.

(3) **Inv-3** holds for k and no event alters buffer contents, so **Inv-3** holds for $k+1$.

[4b] “ $\text{pc}'(t) = \bar{q}_{n-1}$.” Since **[2]** does not hold, $\text{inst}(\mathbf{e}) = \text{inst}_n$. Furthermore, because $\text{count} = \text{max}$, additionally to performing the events that simulate the load behavior as in subcase **[4a]**, the extension returns to the original program flow

using events for (9–10) and makes the auxiliary registers address-value pairs explicit in $\text{buf}'_+(t)$.

Let $\mathbf{e}'_1, \dots, \mathbf{e}'_{\max} \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that \mathbf{e}'_k are, for all $k \in [1..\max]$, the buffering events for stores (9,10), and $s''_{k+1} \xrightarrow{\mathbf{e}'_1 \dots \mathbf{e}'_{\max}} s'_{k+1} \in \Delta_{\text{TSO}}^*$ according to (LS) from Figure 2, with s''_{k+1} being notation for s'_{k+1} from [4a]. We define $\beta_{k+1} := \beta'_{k+1} \cdot \mathbf{e}'_1 \dots \mathbf{e}'_{\max}$, where β'_{k+1} is notation for β_{k+1} from [4a], and find that the invariants **Inv-0...3** hold for $k+1$ since

- (0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.
- (1) **Inv-1** holds for k and $\text{pc}_+(t) = \text{dst}(\text{inst}_n) = \text{pc}'_+(t)$, so **Inv-1** holds for $k+1$.
- (2) **Inv-2** holds for k , both events \mathbf{e} and \mathbf{e}_{load} update r by the same value, and no other event $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_1, \mathbf{e}'_1, \dots, \mathbf{e}'_{\max}$ changes any address, so **Inv-2** holds for $k+1$.
- (3) **Inv-3** holds for k and events $\mathbf{e}'_1, \dots, \mathbf{e}'_{\max}$ place the corresponding address-value pairs that match $\text{buf}_+(t)$ into $\text{buf}'_+(t)$, so **Inv-3** holds for $k+1$.

[4c] “ $\text{pc}'(t) = \text{pc}(t)$.” This case is similar to [3b]. Let $\mathbf{e}' \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that $\text{inst}(\mathbf{e}') = \text{inst}(\mathbf{e})$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ in $X_{\text{TSO}}(R \oplus \sigma)$. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that, by the Δ_{TSO} semantics (Figure 2), the invariants **Inv-0...3** hold for $k+1$.

[5] “ \mathbf{e} performs an assignment, conditional, or memory fence and [2] fails.”

We analyze the following subcases.

[5a] “ $\text{pc}'(t) = \overline{q}_{i-1}$ for $i \in [1..n-1]$.” Since [2] does not hold, $\text{inst}(\mathbf{e}) = \text{inst}_i$ is either a conditional or an assignment.

If $\text{cmd}(\text{inst}_i) = r \leftarrow e$ let $\mathbf{e}' \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that $\text{inst}(\mathbf{e}') = (q_{i-1}, r \leftarrow e, q_i)$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ by the Δ_{TSO} rule for local assignments. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that the invariants **Inv-0...3** hold for $k+1$ since

- (0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.
- (1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{dst}(\text{inst}_i)$, and $\text{pc}'_+(t) = \overline{q}_i$, so **Inv-1** holds for $k+1$.
- (2) **Inv-2** holds for k and e is evaluated the same by both \mathbf{e} and \mathbf{e}' , so the register r is updated by the same value and **Inv-2** holds for $k+1$.
- (3) **Inv-3** holds for k and no event alters buffer contents, so **Inv-3** holds for $k+1$.

Otherwise, $\text{cmd}(\text{inst}_i) = \text{assume } e$. Let $\mathbf{e}' \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that $\text{inst}(\mathbf{e}') = (q_{i-1}, \text{assume } e, q_i)$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ by the Δ_{TSO} rule for conditionals. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that the invariants **Inv-0...3** hold for $k+1$ since

- (0) **Inv-0** holds for k so $s_0 \xrightarrow{\alpha_{k+1}} s_{k+1}$ and $s'_0 \xrightarrow{\beta_{k+1}} s'_{k+1}$, i.e. **Inv-0** holds for $k+1$.
- (1) **Inv-1** holds for k , $\text{pc}_+(t) = \text{dst}(\text{inst}_i)$, and $\text{pc}'_+(t) = \overline{q}_i$, so **Inv-1** holds for $k+1$.

(2) **Inv-2** holds for k and both \mathbf{e} and \mathbf{e}' do not change any address, so **Inv-2** holds for $k + 1$.

(3) **Inv-3** holds for k and no event alters buffer contents, so **Inv-3** holds for $k + 1$.

5b “ $\text{pc}'(t) = \text{pc}(t)$.” This case covers the remaining possibilities when \mathbf{e} is an assignment, conditional, or memory fence. Similar to cases **3b** and **4c**, let $\mathbf{e}' \in E_\oplus$ and $s'_{k+1} \in S_\oplus$ such that $\text{inst}(\mathbf{e}') = \text{inst}(\mathbf{e})$ and $(s'_k, \mathbf{e}', s'_{k+1}) \in \Delta_{\text{TSO}}$ in $X_{\text{TSO}}(R \oplus \sigma)$. We define $\beta_{k+1} := \beta_k \cdot \mathbf{e}'$ and find that, by the Δ_{TSO} semantics (Figure 2), invariants **Inv-0...3** hold for $k + 1$.

The above case distinction covers all possibilities for events \mathbf{e} that α may perform from s_k . Hence, by complete induction, the extension does not remove TSO-reachable states: if $s = (\text{pc}, \text{val}, \text{buf})$ is reachable by α then there exists $s' = (\text{pc}', \text{val}', \text{buf}')$ and $\beta \in \mathcal{C}_{\text{TSO}}(R \oplus \sigma)$ such that s' is reachable by β in $R \oplus \sigma$, $\text{pc} = \text{pc}'$, $\text{val}(a) = \text{val}'(a)$ for all $a \in \text{DOM} \cup \text{REG}$, and $\text{buf} = \text{buf}'$ are empty.

For the reverse direction, let $f_\tau: \mathcal{C}_{\text{TSO}}(R) \rightarrow \mathcal{C}_{\text{TSO}}(R \oplus \tau)$ be the map $\alpha \mapsto \beta$ that the inductive proof implies, respectively $f_\tau: E \rightarrow E_\oplus^*$ its restriction to events matching the different inductive cases. Furthermore, consider computations $\beta \in \mathcal{C}_{\text{TSO}}(R \oplus \sigma)$ that do not interleave events of other threads within the events of sequences $f_\tau(\mathbf{e})$. Such computations reach the entire set $\text{Reach}_{\text{TSO}}(R \oplus \sigma)$. E.g., since local events $\mathbf{e}_{\text{count}}, \dots, \mathbf{e}_1$ as in case **4a** that precede \mathbf{e}_{load} can be performed right before \mathbf{e}_{load} , the above restriction does not change the set of TSO-reachable states in $R \oplus \sigma$. Note that f_τ is a bijection between such computations β and computations $\alpha \in \mathcal{C}_{\text{TSO}}(R)$ that delay flushes locally the least wrt. t . Another induction can show that for each computation β as described above there exists a computation $\alpha \in \mathcal{C}_{\text{TSO}}(R)$ such that invariants **Inv-0...3** hold for prefixes of β and α . This implies that the extension by σ does not add TSO-reachable states. \square

C TSO Semantics and Proofs missing in Section 4

Figure 13 describes the full TSO semantics. For completeness, states $s \in S_M$ use the additional event counter $\text{ec}: \text{TID} \rightarrow \mathbb{N}$ to identify events. This is used, e.g., to define matching stores and flushes and does not affect in any way our results.

As mentioned in subsection 2.1, under SC, stores are flushed immediately:

$$\frac{\text{cmd} = \text{mem}[e_a] \leftarrow e_v, \quad a = \widehat{e}_a, \quad v = \widehat{e}_v, \quad \text{id} = \text{ec}(t)}{s \xrightarrow{(t, \text{id}, \text{inst}, a)(t, \text{id}, \text{flush}, a)} (\text{ec}', \text{pc}', \text{val}[a := v], \text{buf})} \text{ (LSWM)}$$

Lemma 8 *If $\alpha, \beta \in \mathcal{C}_{\text{TSO}}(P)$, $s_0 \xrightarrow{\alpha} s$, and $\rightarrow_{hb}(\alpha) = \rightarrow_{hb}(\beta)$ then $s_0 \xrightarrow{\beta} s$.*

Proof. Assume $s_0 \xrightarrow{\beta} s'$. Since α and β have the same program order \rightarrow_{po} , it means s and s' have the same index counter ec and program counter pc . Moreover, since α and β have the same conflict order \rightarrow_{cf} , s and s' have the same memory valuation val . Finally, since computations α and β empty the buffers, s and s' have empty buffers. In conclusion, $s = s'$. \square

$$\begin{array}{c}
\frac{cmd = r \leftarrow \text{mem}[e_a], \quad a = \widehat{e}_a, \quad \text{buf}(t) \downarrow (\mathbb{N} \times \{a\} \times \text{DOM}) = (id, a, v) \cdot \beta}{s \xrightarrow{(t, \text{ec}(t), \text{inst}, a)} (\text{ec}', \text{pc}', \text{val}[r := v], \text{buf})} \quad (\text{RB}) \\
\\
\frac{cmd = r \leftarrow \text{mem}[e_a], \quad a = \widehat{e}_a, \quad \text{buf}(t) \downarrow (\mathbb{N} \times \{a\} \times \text{DOM}) = \varepsilon, \quad v = \text{val}(a)}{s \xrightarrow{(t, \text{ec}(t), \text{inst}, a)} (\text{ec}', \text{pc}', \text{val}[r := v], \text{buf})} \quad (\text{RM}) \\
\\
\frac{cmd = \text{mem}[e_a] \leftarrow e_v, \quad a = \widehat{e}_a, \quad v = \widehat{e}_v, \quad id = \text{ec}(t)}{s \xrightarrow{(t, id, \text{inst}, a)} (\text{ec}', \text{pc}', \text{val}, \text{buf}[t := (id, a, v) \cdot \text{buf}(t)])} \quad (\text{LS}) \\
\\
\frac{\text{buf}(t) = \beta \cdot (id, a, v)}{s \xrightarrow{(t, id, \text{flush}, a)} (\text{ec}, \text{pc}, \text{val}[a := v], \text{buf}[t := \beta])} \quad (\text{WM}) \quad \frac{cmd = \text{mfence}, \quad \text{buf}(t) = \varepsilon}{s \xrightarrow{(t, \text{ec}(t), \text{inst}, \perp)} (\text{ec}', \text{pc}', \text{val}, \text{buf})} \quad (\text{LF}) \\
\\
\frac{cmd = r \leftarrow e, \quad v = \widehat{e}}{s \xrightarrow{(t, \text{ec}(t), \text{inst}, \perp)} (\text{ec}', \text{pc}', \text{val}[r := v], \text{buf})} \quad (\text{LA}) \quad \frac{cmd = \text{assume } e, \quad \widehat{e} \neq 0}{s \xrightarrow{(t, \text{ec}(t), \text{inst}, \perp)} (\text{ec}', \text{pc}', \text{val}, \text{buf})} \quad (\text{LC})
\end{array}$$

Fig. 13. Transition rules for $X_{\text{TSO}}(P)$ assuming $s = (\text{ec}, \text{pc}, \text{val}, \text{buf})$ with $\text{pc}(t) = q$, $\text{inst} = q \xrightarrow{cmd} q'$ in thread t , $\text{ec}' = \text{ec}[t := \text{ec}(t) + 1]$, $\text{pc}' = \text{pc}[t := q']$. We use \widehat{e} to evaluate e under val and $\text{buf}(t) \downarrow (\mathbb{N} \times \{a\} \times \text{DOM})$ for stores in $\text{buf}(t)$ that access a .